Scale-free law: network science and copyright

Article (Accepted Version)

SCALE-FREE LAW: NETWORK SCIENCE AND COPYRIGHT

*Andrés Guadamuz González**

I. INTRODUCTION

One common feature of literature dealing with new technologies, and particularly those authors dealing with any legal aspect of cyberspace, is to over-emphasize the importance of information and communication technologies to our present situation. This has led to an abundance of commonplaces that are just variations of the phrase "the Internet has changed everything." While it is useful to distance oneself from such clichés, this essay will unfortunately begin with a variation of this theme. Commonplaces exist for a reason, and in the areas of copyright law and the study of networks, the Internet has indeed changed everything.

Networks are everywhere. The staggering complexity and seemingly chaotic nature of everyday life is actually a collection of different interactions. We are constantly surrounded by the social network, the financial network, the transport network, the telecommunications network, and even the network within our own bodies. The understanding of how these systems operate and interact with one another has been the realm of physicists, economists, biologists, and mathematicians. Until recently, the study of networks was left to theoretical and academic debates as it lacked proper empirical application because it was difficult to gather reliable data about large and complex systems. But in recent years, the Internet has given researchers the opportunity to study and test the mathematical descriptions of the vast complex systems. The growth rate and structure of cyberspace allows researchers to map and test several previously unproven theories about how links and hubs within the network interact with one another. With the Web, we now have means to test the organizational structures of networks, their architecture, their growth, and even allow some limited predictions about their behavior, strengths and vulnerabilities.

With the increasing reliability on the descriptive—and sometimes predictive—nature of network science, a logical next step for legal scholars is to look at the potential legal implications of some of the characteristics of networks. Some academics and practitioners have started finding potential uses for network science tools; efforts that will be highlighted in following pages. One particular topic of interest where network science could have a noticeable effect is the area of the regulation of the Internet, which has provided ample possibilities for discussion and analysis during its short existence.[1] Some of the most interesting legal literature from the early days of the modern Internet[2] deals with the potential difficulties in putting

* Lecturer, University of Edinburgh, and co-director of the AHRC Research Centre for Studies in Intellectual Property and Technology Law.

[1] Some of these arguments are masterfully brought together in JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET? ILLUSIONS OF A BORDERLESS WORLD (2006).

[2] The author has chosen the arbitrary date for the birth of the "modern" Internet as of December 15, 1994, when Netscape Communications launched its graphical browser called Mozilla 1.0, giving rise to the graphical web. *See* Netscape—The First Commercial Web

a leash on the chaotic and anarchic nature of cyberspace. The skepticism about the impossibility to generate any effective type of regulation has prompted some authors to theorize about how to exercise control over the online world.[3] This debate left some unanswered questions, particularly in the area of intellectual property rights, where the enforcement of copyright in the digital domain has become an increasingly difficult issue.

It is in this legal context where the present article will attempt to make use of the study of self-organized scale-free networks in order to make use of the improved understanding of how networks are formed, grow and operate, and apply it to problematic regulation of copyright in cyberspace.

II. THE NETWORK SCIENCE REVOLUTION

It would be easy to overestimate the importance of network theory in the real world by stressing the importance of the Internet to our everyday lives; however, such assumptions would be missing the fact that the study of networks is not a new science. The understanding of how networks operate and interact with one another has been studied by physicists, economists, and mathematicians for centuries.[4] Many operational assumptions and theories of networks, however, had not been applied by other disciplines. Some earlier works on the topic described specific network architectures and characteristics, but the studies were smaller-scale, although they set the theoretical principles for what was to become the modern discipline.[5]

In recent years, the Internet has given researchers the opportunity to study and test several of the existing mathematical descriptions of complex networks.[6] Although the Web is composed of billions of pages, its fast growth-rate and international reach allow researchers to map and examine several previously untested ideas about how networks interact. With a combination of the characteristics of online hyper-linking, and the help of spiders and web crawlers,[7] researchers have the means to test the organizational structures of the architecture and behavior of networks.

Much of the current interest in networks can be traced back to a series of popular science books dedicated to publicizing the latest

---

Browser, Marc Andreessen, Jim Clark, Mosaic, http://www.livinginternet.com/w/wi_netscape.htm (last visited Mar. 29, 2007).

[3] For example, some authors have examined the idea of "Net Federalism." *See generally* LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE 85–99 (1999) (discussing the threats to liberty in cyberspace); David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1367–80 (1996) (discussing the parameters of the law in cyberspace); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998) (discussing the modern influence of Lex Informatica).

[4] Leonhard Euler's classic paper *Seven Bridges of Königsberg* was published in 1736. *See* ScienceWeek, History of Mathematics: On Leonhard Euler (1707–1783), http://scienceweek.com/2003/sc031121-6.htm (last visited Feb. 12, 2007).

[5] A pivotal work in early graph theory (the precursor of network science) is P. Erdös & A. Rényi, *On the Evolution of Random Graphs*, 38 BULL. DE L'INSTITUT INTERNATIONAL DE STATISTIQUE 343 (1961).

[6] *See, e.g.*, Michalis Faloutsos et al., *On Power-Law Relationships of the Internet Topology*, 29 APPLICATIONS TECH. ARCHITECTURES PROTOCOLS COMPUTER COMM. 251, 251 (1999) (discovering simple power-laws of the Internet); Andrei Broder et al., *Graph Structure in the Web*, *available at* http://www.cindoc.csic.es/cybermetrics/pdf/22.pdf (last visited Mar. 29, 2007) (reporting on experiments on the web).

[7] A web crawler is a computer program that browses the Internet in an automated and predetermined manner. *See* Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (1998), *available at* http://infolab.stanford.edu/pub/papers/google.pdf.

developments in research. Titles of note are *Linked* by Albert-László Barabási,[8] *The Tipping Point* by Malcom Gladwell,[9] *Critical Mass* by Philip Ball,[10] and *Six Degrees* by Duncan J. Watts.[11] These "pop science" credentials could make those unfamiliar with the literature suspicious about the validity and reliability of network theories,[12] but this skepticism would be misplaced, as most of these books have sound peer-reviewed research behind them, and in most instances they have been written by the primary investigators themselves.

The modern understanding of networks begins with the study of statistical phenomena called power laws, described as "[w]hen the probability of measuring a particular value of some quantity varies inversely as a power of that value."[13] In other words, power laws are tools that describe the divergence in the predictable and average value of an observable fact. Most natural phenomena display "normal" distributions, which when plotted in an axis display a bell-shaped form. In a normal distribution, the largest number of instances is encountered in the middle. Most people are average height, while there are small numbers of both very short and very large people; charting such distribution will provide a bell-shaped curve.[14]

Power distributions, however, do not follow the normal trend; in them we find that there are few remarkable instances that account for a very large number of occurrences of the studied event. Because of this, a power law distribution does not have a peak, but the line charting given incidences of an event tend to drop off sharply, which indicates the increased likelihood of extreme events.[15] An example of power law distributions can be found in city populations. If we are counting all of the people living in cities around the world, we will soon discover that megalopolis like Tokyo, Mexico City, New York, and Mumbai account for a disproportionate amount of the total city inhabitants. These cities generate tell-tale spikes in the data, accompanied by a long tail of smaller populations.

Power laws often display what is known as Pareto distributions,[16] or the 80/20 rule, following the popular perception that 80 percent of the work is performed by 20 percent of the employees; or that 80 percent of the wealth is held by 20 percent of the population.[17] A Pareto distribution, named after economist Vilfredo Pareto, is used to describe large inequalities in data, where most of the distribution is concentrated in a relatively small portion of the graph.

It may be surprising that power laws seem to be found in all sorts

---

[8] ALBERT-LÁSZLÓ BARABÁSI, LINKED: THE NEW SCIENCE OF NETWORKS (2002).

[9] MALCOLM GLADWELL, THE TIPPING POINT: HOW LITTLE THINGS CAN MAKE A BIG DIFFERENCE (2002).

[10] PHILIP BALL, CRITICAL MASS: HOW ONE THING LEADS TO ANOTHER (2004).

[11] DUNCAN J. WATTS, SIX DEGREES: THE SCIENCE OF A CONNECTED AGE (2002).

[12] It should be noted that network theory should not be confused with actor-network theory. For the difference between the two, see BRUNO LATOUR, REASSEMBLING THE SOCIAL: AN INTRODUCTION TO ACTOR-NETWORK-THEORY (2005). McLuhan also has something to say about networks, and is often cited as the father of network theory. *See* PAUL LEVINSON, DIGITAL MCLUHAN: A GUIDE TO THE INFORMATION MILLENNIUM 187–200 (2001) (discussing McLuhan).

[13] M.E.J. Newman, *Power Laws, Pareto Distributions and Zipf's Law*, 46 CONTEMP. PHYSICS 323, 323 (2005).

[14] *See generally* STEPHEN M. STIGLER, STATISTICS ON THE TABLE: THE HISTORY OF STATISTICAL CONCEPTS AND METHODS 403–15 (1999) (discussing the concept of "normal" in relation to the bell shaped curve).

[15] *See* BALL, *supra* note 10, at 104–07.

[16] William J. Reed, *The Pareto, Zipf and Other Power Laws* 1–2, *available at* http://linkage.rockefeller.edu/wli/zipf/reed01_el.pdf (last visited Mar. 29, 2007).

[17] *See* BARABÁSI, *supra* note 8, at 66.

of circumstances outside of what is normally perceived as a network, including biological systems.[18] Other places where these networks have been found are, according to Newman:

> In addition to city populations, the sizes of earthquakes, moon craters, solar flares, computer files and wars, the frequency of use of words in any human language, the frequency of occurrence of personal names in most cultures, the numbers of papers scientists write, the number of citations received by papers, the number of hits on web pages, the sales of books, music recordings and almost every other branded commodity, the numbers of species in biological taxa, people's annual incomes and a host of other variables all follow power-law distributions.[19]

A corollary of power laws is that this type of distribution may produce scale-free environments. In a normal distribution, there is little or no room for results that are considerably above and below the norm. In a plot of people's heights, one will expect to find that most people are average, with deviations towards both ends, thus forming a bell when charted. In a scale-free environment, most people would be average height, while there would be some thirty and fifty meter giants walking around, and from time to time you could even encounter a person measuring hundreds of meters.[20]

Power laws and scale-free topologies apply to networks in general, but they seem to be more prevalent in large-scale complex systems.[21] In order to chart vast networks, one must understand some of the basics of how they operate. Most networks are composed of three elements: *nodes* are individual elements in the network; *links* are the relations between nodes; while *hubs* are collections or clusters of nodes.[22] In a normal network distribution, we would expect to find that nodes are distributed in an average manner, some with more links, and some with fewer links; this generates a random chart. In a scale-free network, the vast majority of nodes and hubs have an average or small number of links, while very few hubs will have an exceptionally large number of links, forming super-nodes, or even super-hubs.[23]

Thanks to the wealth of innovative research into networks, the architecture of the Internet is now understood enough to claim that it presents many of the inherent characteristics of scale-free networks and, therefore, it can be said that it responds to power laws. The topology of the Internet has allowed its study by providing researchers with hyperlinks, which are a ready-made tool for measuring connectedness. Spiders and other autonomous agents can be programmed to trawl the Web in order to gather information about the pages, sites and links. This has allowed researchers to confirm the features of the Internet and understand its underlying architecture with an amazing degree of certainty.[24]

This ability has in turn allowed the charting of certain laws of the Internet. Amidst the seemingly chaotic nature of the Web, a hidden

---

[18] H. Jeong et al., *The Large-Scale Organization of Metabolic Networks*, 407 NATURE 651, 651 (2000).

[19] Newman, *supra* note 13, at 325 (internal citations omitted).

[20] BARABÁSI, *supra* note 8, at 67–69.

[21] Erzsebet Ravasz & Albert-László Barabási, *Hierarchical Organization in Complex Networks*, 67 PHYSICAL REV. E 026112, 026112-1 (2003).

[22] *Id.*

[23] BARABÁSI, *supra* note 8, at 69–72.

[24] For more about this, see BERNARDO A. HUBERMAN, THE LAWS OF THE WEB: PATTERNS IN THE ECOLOGY OF INFORMATION 24, 30 (2001).

regularity emerges in every studied pattern. For example, web sites under a domain seem to respond to power laws in the way in which pages are visited. The hub tends to be the home page, and subsequent links from the main site tend to decrease markedly into a power law distribution.[25] Similarly, web site popularity displays considerably few popular pages, with sharp drop offs into a long tail of less visited sites.[26] The resulting clustering tends to produce an ecology dominated by hubs and super-hubs that act as the glue that binds and controls web traffic. This would result in the Internet not being at all a random space, as the likelihood for an average user to visit a web site responds to power laws.[27] One of the main features of the Internet is that its growth responds to the expected accumulation of links, which is one of the trademarks of scale-free networks. Few websites accumulate staggering numbers of links, while the vast majority of sites have few links, which constitute a textbook example of a power-law.[28] Not only is there a power law at work in cyberspace, but the rate of accumulation of sites responds to how long they have been accumulating links, which serves to confirm its scale-free architecture.[29] This can be seen in the manner in which websites like Google and Yahoo act as hubs in the Web landscape.

The clustering described above explains one of the most publicised corollaries arising from the research into scale-free networks, and that is the phenomenon of so-called six degrees of separation, or small world distributions. This is a commonly-held knowledge that all of the people in the world are separated only by six connections from one another. This belief comes from a study by psychologist Stanley Milgram, who tried to measure how many links there were between sixty people in Kansas and one target in Massachusetts, which resulted in a surprisingly small number of intervening connectors.[30]

When the small world phenomenon has been put to the test, it has proven to be surprisingly accurate,[31] although social networks seem to be starkly divided by economic and ethnic sub-networks.[32] The reason why there is a correlation between this hypothesis and scale-free systems is evident if one considers that there are certain hubs in social networks that acquire more links than others. These hubs act as “connectors”,[33] and, once one has reached a connector, chances are that it will provide a large number of links to other nodes in the system. As the Internet is a scale-free network, popular sites will act as hubs, allowing large interconnectedness between nodes, demonstrating once more the reliability of the small world distribution. According to research into average node length online by mapping links, the average path between two random websites is as small as 4.22 links.[34]

Another interesting characteristic of scale-free networks is that

---

[25] *Id.* at 30.

[26] *Id.* at 46–49.

[27] *Id.* at 23–25.

[28] Réka Albert et al., *Diameter of the World-Wide Web*, 401 NATURE 130, 130 (1999).

[29] Soon-Hyung Yook et al., *Modeling the Internet's Large-Scale Topology*, 99 PROC. NAT’L ACAD. SCI. 13382–86 (2002).

[30] *See* Stanley Milgram, *The Small-World Problem*, PSYCHOL. TODAY, May 1967, at 60–67.

[31] *See* Richard J. Williams et al., *Two Degrees of Separation in Complex Food Webs*, 99 PROCEEDINGS NAT’L ACAD. SCI. 12913, 12913, 12915 (2002) (finding, on average, two degrees of separation between species in complex food webs).

[32] *See* Judith S. Kleinfeld, *The Small World Problem*, 39 SOCIETY 61, 65 (2002).

[33] GLADWELL, *supra* note 9, at 34–64.

[34] Huberman discusses an experiment that obtained the average number of links from among 64,826 sites. *See* HUBERMAN *supra* note 24, at 37–38.

they tend to display remarkable robustness and stability.[35] Strogatz explains that "scale-free networks are resistant to random failures because a few hubs dominate their topology. . . . Any node that fails probably has small degree (like most nodes) and so is expendable."[36] In other words, if one tries to attack a scale-free network randomly, the result will be that the attacked node will be unlikely to play any importance in the way in which the network stays together. This is because hubs tend to be few, so the chances of hitting one randomly are very high. The Internet has proved to have inherited such robustness,[37] as virus attacks, and even Distributed Denial of Service (DDoS)[38] have not managed to bring the entire network down.

While scale-free networks are strong, they are not invulnerable. There are documented circumstances where scale-free systems have collapsed in spectacular fashion due to cascading failures. In 1996, a large blackout affected eleven states in the United States and two Canadian provinces, which originated from the failure of one single line in Oregon.[39] Energy grids are typical examples of scale-free networks because they rely on a few key hubs in order to maintain distribution loads. If one of those hubs is removed, the entire system may collapse, an effect that spells the vulnerability of networks to random occurrences in hubs,[40] or even to targeted attacks against one.[41]

The way in which scale-free networks operate presents a number of other effects besides the power law distribution of nodes and links described. One such corollary arising from the study of networks is that as a network grows, popular nodes and hubs will continue to gather more links as time goes by:[42] an effect also known as the rich get richer. This effect takes place because of the cumulative effect of the interaction between the links. The older a node is, the more likely it will be to have established links, and to have been communicated to other nodes, while newer nodes will lack this advantage. The accumulation of links can lead to a collapse of node competition, and one node becoming the sole super-hub, a phenomenon known as the "winner-takes-all."[43]

All of these seeming deterministic descriptions of complex systems such as the Internet seem too good to be true. How is it possible for a seemingly chaotic system to become organized and display behavior that can be predicted with such ease? It seems counter-intuitive to expect a complex environment to display self-organization when one would expect the contrary. The science of phase transitions explains this. Phase transitions take place when

---

[35] *See* Réka Albert et al., *Error and Attack Tolerance of Complex Networks*, 406 NATURE 378, 378 (2000) ("Many complex systems display a surprising degree of tolerance against errors.").

[36] Stevon H. Strogatz, *Exploring Complex Networks*, 410 NATURE 269, 274 (2001).

[37] Yuhai Tu, *How Robust is the Internet?*, 406 NATURE 353, 353–54 (2000).

[38] A DDoS is an attack on a computer network by more than one system "that causes a loss of service to users, typically the loss of network connectivity and services by consuming the bandwidth of the victim." Lilian Edwards, *Dawn of the Death of Distributed Denial of Service: How to Kill Zombies*, 24 CARDOZO ARTS & ENT. L.J. 23, 23 (2006) (citation omitted).

[39] BARABÁSI, *supra* note 8, at 119.

[40] *See* Y. Moreno et al., *Instability of Scale-Free Networks Under Node-Breaking Avalanches*, 58 EUROPHYSICS LETTERS 630, 630 (2002) (examining how the failure of one node in a complex network can lead to subsequent failures, thereby contributing to the instability of the network).

[41] *See* Béla Bollobás & Oliver Riordan, *Robustness and Vulnerability of Scale-Free Random Graphs*, 1 INTERNET MATHEMATICS 1, 5 (2003).

[42] *See* P. L. Krapivsky et al., *Degree Distributions of Growing Networks*, 86 PHYSICAL REV. LETTERS 5401, 5401 (2001).

[43] Ginestra Bianconi & Albert-Lászlo Barabási, *Bose-Einstein Condensation in Complex Networks*, 86 PHYSICAL REV. LETTERS 5632, 5633–34 (2001).

a chaotic network enters a meta-stable transitional period after which the system becomes ordered rapidly.[44] Scale-free systems tend to display such behavior because hubs move the system into a specific direction.

The result of all of these characteristics and corollaries is that the Internet has become a space much easier to chart than was previously expected. Websites accumulate links in accordance to well-established rules. The Web's ecology of pages displays a certain order in clusters united by super-hubs. The average number of links from one page to another allows for the small world phenomenon. To paraphrase Huberman, the Internet has a set of Laws that determines its architecture.

III. NETWORK SCIENCE AND SOCIETY

A. The Deterministic Net?

The network science research described above may generate unease amongst some readers.[45] This may be because the use of physical formula to understand human behaviour has had a bad history. These attempts are reminiscent of prior efforts to marry physics and society, exemplified by Hobbes' *Leviathan*, where the philosopher wondered about the existence of physical principles governing celestial bodies, and thought that similar principles might exist to explain the interaction of social agents.[46] The implication of such a deterministic outlook of the world has had negative implications even for Hobbes, so it has to be suspected by default. But despite its dubious history, modern physics has been demonstrating that there could be an application of physical models to social interactions.[47] Formulas used to describe how magnets achieve their orientation, or how gases condense, can also be used to chart how businesses grow, how crime rates fluctuate, or how crowds flow.[48]

It would be easy to dismiss the trends cited, and particular the emerging science of networks, as another doomed attempt to explain social complexity with mathematics, or a way of deleting free will to convert the human experience into a set of equations. To view power laws as deterministic, however, does not really address the fact that this is not an exact science; it is a descriptive tool of how networks operate.[49] Humans still retain free agency, while the network itself could be deterministic and react in predicable ways.

The best way to understand the potential deterministic nature of networks is to conduct simple thought experiments about how people actually interact with one another in a social gathering. We would generally like to think that we are free agents, and therefore social networks should respond to the very random nature of human

---

[44] *See* BALL, *supra* note 10, at 100–03; *see also* STEVEN H. STROGATZ, SYNC: HOW ORDER EMERGES FROM CHAOS IN THE UNIVERSE, NATURE, AND DAILY LIFE 230, 230–36 (2003) (discussing plans and actions taken to test for phase transitions).

[45] If the author's experience at some conferences can be used as anecdotal measure of such feeling, then there might be plenty of animosity towards the application of network science to social systems.

[46] *See* THOMAS HOBBES, HOBBES'S LEVIATHAN 82 (Oxford Univ. Press 1929) (1651).

[47] *See generally* Philip Ball, *The Physical Modelling of Human Social Systems*, 1 COMPLEXUS 190 (2003) (analyzing statistical physics to explain human social interactions).

[48] *Id.* at 198–200.

[49] It must be stressed that the term "deterministic nature" has other implications in the research. It is another mathematical model to describe network growth. *See generally* Albert-László Barabási et al., *Deterministic Scale-Free Networks*, 299 PHYSICA A 559, 559–60 (2001) (proposing a scale-free network model in a deterministic fashion).

experience. Yet, we are constantly responding and acting according to physical and social constraints. Imagine that you are at a busy conference coffee break. If you are observant, you will probably notice that people have gathered in small groups, some people will work the room, while others will remain with the same group, and perhaps there will even a person standing by the coffee table on his or her own. You will rarely see a person shouting across the room, or an extremely large group where nobody can interact. If you map the number of links made during such breaks, you will start to see certain patterns emerging. These patterns are not deterministic in the sense that they completely erase agency from those present; you can still choose to move around the room, or not to talk to anyone else; however, the pattern made by the collection of conducts provides a good example of the apparently deterministic nature of social networks. People will act freely, but the constraints of social norms and the laws of physics will mean that social networks will produce certain results. Smaller groups will have less deterministic value because the action of one individual will have a larger effect, while the larger group will tend to absorb the random individual behavior.[50] This same phenomenon is precisely what has been mapped by the research conducted so far on all sorts of networks.[51] Large scale-free networks seem to follow certain rules that respond to those same physical constraints. Meaningful links, nodes, and hubs serve to explain the larger picture, but not the individual choices.

The Internet behaves in this seemingly deterministic way. From the emerging research into Web behavior, it has become evident that links, nodes, and hubs operate in a certain manner.[52] The Web responds to a set of rules. Older links are more likely to have more links than newer ones. Some well-designed computer viruses propagate while others fail to attack. The network itself is robust. The study of large networks such as the Internet does not and cannot study individual behavior; its objective is to understand the "behavior of the system as a whole."[53] The architecture shapes the behavior of the entire network, just like gravity shapes how we interact with the world.

It is only natural that grand theories of everything should be met with skepticism. Attempting to explain complex systems with a few theories may seem like unforgivable reductionism, an attempt to apply materialistic ideals to social relations where they do not fit. If there is sound evidence, however, that certain network environments like the Internet act in predictable ways, then all the research into this behavior should be taken into consideration when attempting to analyze the underlying trends that govern such patterns, even if it is an analysis that belongs to the physical sciences and not to the social ones.

It can be argued that we are on the threshold of better understanding complex systems like the Web thanks to the predictable nature of the science of networks, but it is important to make sure that such enthusiasm is tempered by the scale of the task of mapping such large structures. All predictive models of cyberspace should take into consideration that it is a changing

[50] BALL, *supra* note 10, at 458–62.

[51] *See* BARABÁSI, *supra* note 8, at 145 fig.11.1 (depicting Paul Baran's three possible network designs: centralized, decentralized, and distributed).

[52] To stress the wealth of the research into models describing the Internet, see Z. Dezso et al., *Dynamics of Information Access on the Web*, 73 PHYSICAL REV. E 066132, 066132-2 (2006).

[53] HUBERMAN, *supra* note 24, at 23.

environment. As Barabási argues:

> It is far from us to suggest that the scale-free model introduced above describes faithfully the topology of the www. Naturally, the www has a much richer structure, that cannot be captured by such simple ingredients. For example, the links are not invariant in time, they constantly change, being either eliminated or rewired to other documents. Similarly, the www documents are not stable, they are often removed, and change address. Furthermore, the web pages are structured in domains, that by themselves have a rather complex hierarchical structure.[54]

Research into networks should then be released with the caveat that the descriptive and predictive analysis given to us by studies into power laws and network are to be taken as tools, not as absolute predictions. This has to be stressed because it would be plausible to read the extensive research presented so far and complain that we are talking about a form of technological determinism.[55] The reader can rest assured that such a goal is not intended, and that the tools put forward should not be construed as deterministic in any way, just like gravity is not deterministic.

B. Network Law

The legal reader who has made it this far may be justified in asking the question of what it all means for the legal profession and research. Networks obeying certain rules and presenting specific architectures may be interesting to physicists, not to lawyers. At the time of writing, legal scholarship regarding the interaction between scale-free networks and the law has been scarce, but there are a few exceptions. It is possible that the mathematics and the technical jargon of some of the papers may have dissuaded more interest in the topic. Nevertheless, if some better understanding of networks has been made possible by emerging theories, then the law should take interest in the subject to ascertain if there may be some legal issues worth exploring.

Perhaps one of the most evident areas of study with regard to networks may very well be the regulatory arena. If we can understand a specific network that has given problems to regulators, then the potential for empirical-based research on how the network operates could provide clues as to how to regulate the troublesome area. A study has already attempted to look into the application of specific network theories to the telecommunications field.[56] Recognizing that telecommunication networks operate as a complex system, the authors state that the specific graphical representation of networks into hubs and nodes may be of use in trying to regulate emerging technologies such as access to broadband services and Voice-over-IP (VoIP) communications.[57] This study has a narrow objective, as it relies only on the describing

---

[54] Albert-László Barabási et al., *Scale-Free Characteristics of Random Networks: The Topology of the World-Wide Web*, 281 PHYSICA A 69, 75 (2000).

[55] For more about the topic of technological determinism, see generally DOES TECHNOLOGY DRIVE HISTORY? THE DILEMMA OF TECHNOLOGICAL DETERMINISM (Merritt Roe Smith & Leo Marx eds., 1994).

[56] *See* Daniel F. Spulber & Christopher S. Yoo, *On the Regulation of Networks as Complex Systems: A Graph Theory Approach*, 99 NW. U. L. REV. 1687, 1691–93 (2005).

[57] A complex system is "a system in which its elements interact in ways that transcend any organizing principles being applied to the network, allowing the network to evolve and adapt to environmental changes." *Id.* at 1694 (citation omitted).

power of network science in order to provide regulators with the basis for charging over communications in complex telecoms networks. Is there room for a wider area of application?

Strahilevitz offers a more general approach by researching the legal implications of power laws and scale-free topographies in a ground-breaking analysis of the potential use of network science to the protection of privacy.[58] He uses the specific application of social network theories, such as small world distributions to conclude that the scale-free nature of some social networks may provide us with tools with which we can measure the number of acquaintances that a member of the social system is likely to have. Then he proposes the fact that an individual involved in tort disputes about personal privacy may have the evidentiary means to measure the potential damage to his or her reputation and, therefore, a judge would be able to discern if there has been some actual damage done. He comments that:

> In a tort suit, courts are always called upon to examine causation: would the plaintiff have been harmed in the absence of the defendant's actions? Social networks theory provides a basis for evaluating that question when the plaintiff's injury stems from dissemination of previously private information. Courts simply need to ask themselves: was the widespread dissemination of this information inevitable, or did the defendant's actions materially affect the extent of subsequent disclosure?[59]

This is an elegant use of existing theories in order to provide a direct causal relationship to establish damages; however, one may be wary of establishing the causal link in the first place. If there is one thing that we have learned it is that scale-free networks predict that there will be super-connected nodes in a social network,[60] individuals whose social interaction exceeds by various degrees the average. The person involved in the dispute could very well be one of those, and the calculation of actual damage could prove to be uncertain.

Another potentially valuable application of network theories in the law is in environmental policy-making. The life-sciences have had extensive experience in the use of empirical data in order to design policy in environmental and public health fields.[61] The better understanding of complex environmental systems brought by some of the literature could be used in assessing risks posed by environmental threats, real or imagined.[62] Farber explains the use of power laws to design methods for assessing risks:

> The presence of statistical power laws supports the use of conservative methods of assessing risk. To be more specific, suppose that we are designing a

[58] Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919, 919–23 (2005).

[59] *Id.* at 975.

[60] *See, e.g.*, Ithiel de Sola Pool & Manfred Kochen, *Contacts and Influence*, *in* THE SMALL WORLD 3–16 (Manfred Kochen ed., 1989) (explaining the "small world problem").

[61] Or not. *See, e.g.*, CHRIS MOONEY, THE REPUBLICAN WAR ON SCIENCE 248 (2005) ("The politicization of science presents a severe challenge to modern democratic governments, which depend on a creative tension between elected representatives on the one hand, and unelected technocratic elites on the other.").

[62] *See generally* Daniel A. Farber, *Probabilities Behaving Badly: Complexity Theory and Environmental Uncertainty*, 37 U.C. DAVIS L. REV. 145, 156–61 (2003) (examining risk assessment). For a less successful yet interesting attempt at marrying biotechnology and network science, see Jim Chen, *Webs of Life: Biodiversity Conservation as a Species of Information Policy*, 89 IOWA L. REV. 495 (2004).

> procedure to identify any proposal posing a significant risk, with significance defined as some specific risk level such as one in ten thousand. . . . The only assumption is that among the relevant set of proposals, harmful effects follow a power-law distribution. If so, conservative test procedures may be warranted.[63]

In other words, in a scale-free environment we may expect harmful effects to occur, which are considerably higher than the average witnessed occurrences. If empirical research points towards the existence of power law distributions in a phenomenon that requires regulation, then conservative policies should be followed. This could certainly be useful if one considers that hurricanes appear to display scale-free characteristics.[64] Similar precautionary approaches could be taken in other life-science fields, particularly in public health policy. Pandemics like AIDS seem to follow scale-free behaviors,[65] where a few individuals can infect large numbers of people in a community by their role as connectors.[66] Public policy towards social pandemics like sexually transmitted diseases could be designed to look for these hubs and attempt to treat them first.[67]

The possibility of following links and clusters of nodes and hubs means that the descriptive power of network science can be easily tested in fields with pre-existing network-like characteristics. In the wider network research, a popular experimentation tool has been to chart citation between authors, or to play small world problems with co-authorship networks.[68] Such characteristic is extremely useful in patent law, where a tool that analyzes the cross-citation of previously issued patents would be of extreme use for examiners and inventors. Strandburg has written an excellent study looking at the clustering of cites in patents issued by the United States Patent and Trademarks Office (USPTO), which has demonstrated, amongst other things, that there seems to be increasing stratification in patent cite-ability since the 1980's.[69] This means that a few patents are being cited with more frequency than in the past. Strandburg argues that this could be correlated with decreasing patent quality[70] experienced in the corresponding period. Another very interesting avenue of research explored in this paper is the possibility of an improved manner in which to classify patent claims. Currently, patent subject matter is assigned by examiners in an ad hoc fashion. Strandburg suggests that citation of previous patents may help in assigning the claim to a cluster, which would make its identification much easier.

Another field that could benefit from the better understanding of

[63] Farber, *supra* note 62, at 160.

[64] SURAJE DESSAI & MARTIN E. WALTER, SELF-ORGANISED CRITICALITY AND THE ATMOSPHERIC SCIENCES: SELECTED REVIEW, NEW FINDINGS AND FUTURE DIRECTIONS 1 (2000), *available at* http://www.isse.ucar.edu/extremes/papers/walter.PDF.

[65] Zoltán Dezső & Albert-Lászó Barabási, *Halting Viruses in Scale-Free Networks*, 65 PHYSICAL REV. E 055103, 055103-1 (2002).

[66] An example of this is the so-called patient-zero of the AIDS pandemic. *See generally* RANDY SHILTS, AND THE BAND PLAYED ON: POLITICS, PEOPLE, AND THE AIDS EPIDEMIC (1987) (discussing the AIDS pandemic).

[67] *See* Ian Ayres & Katharine K. Baker, *A Separate Crime of Reckless Sex*, 72 U. CHI. L. REV. 599, 610–14 (2005).

[68] S. Redner, *How Popular is Your Paper? An Empirical Study of the Citation Distribution*, 4 EUR. PHYSICAL J. 131, 131 (1998).

[69] KATHERINE J. STRANDBURG, LAW AND THE SCIENCE OF NETWORKS: AN OVERVIEW AND AN APPLICATION TO THE 'PATENT EXPLOSION' 54 (2006), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=926354#PaperDownload.

[70] For more about patent quality, see generally ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS 60–69 (2004).

how networks operate and organize is legal scholarship. As cited above, studies have demonstrated that co-authorship and scientific research respond to power laws,[71] with the topology presented as close-knit clusters of researchers.[72] Similarly, there is an interesting paper looking at the way in which the United States Supreme Court decisions cite one other.[73] The study found that there is a scale-free topology at work as there are some decisions that are cited with disproportionate frequency. According to the study, the cases that act as the most cited hubs in this network of citations are older decisions regarding Federal Jurisdiction.[74] It could be said that such a study may not be particularly enlightening, as it does not really say much about the actual nature of the rulings, but similar exercises could be of use for constitutional lawyers in all jurisdictions in order to recognize which cases they are more likely to encounter in future decisions.

A much more controversial subject could be the use of power laws in areas such as criminology and law enforcement. Barabási makes an impassioned yet unconvincing argument of the potential use of network science in the detection of terrorist cells.[75] He explains that the understanding of social networks could be used in the "War on Terror" against enemy cells, as they could be vulnerable to targeted attacks against the hubs holding together the network. Unfortunately, this seems fabricated and rests on completely untested assumptions, unlike the rest of his work.[76] It is a tantalizing promise, but one would need evidence that terrorist and criminal organizations behave according to power laws.

IV. COPYRIGHT LAW AND NETWORKS

The many examples cited in the previous section provide ample evidence of the potential uses of network science to legal scholarship, policymaking and practice. However, at the time of writing there has not been a specific suggestion in the literature about tying this research to copyright law.[77] A lot of effort has gone into attempting to regulate the global network, and copyright has been at the forefront of such efforts. Could the understanding of power laws in the Web provide some solutions to the regulation of the Internet?

A. Enforcement: Peer-to-Peer Networks

It has been a starting point of the present article that the regulation of copyright in online environments has been extremely difficult. After the emergence of the modern Internet, it became evident that the new environment had become a perfect environment for indiscriminate copying of works.[78] The

[71] *See generally* A.L. Barabási et al., *Evolution of the Social Network of Scientific Collaborations,* 311 PHYSICA A 590 (2002).

[72] Perhaps this could be an empirical manner of applying for funding?

[73] Seth J. Chandler, *The Network Structure of Supreme Court Jurisprudence* 1 (Univ. of Houston Law Ctr., Paper No. 1, 2005), *available at* http://ssrn.com/abstract=742065.

[74] *Id.* at 18 (stating that the top two cited cases are *M'Culloch v. Maryland*, 17 U.S. (4 Wheat.) 316 (1819) and *Gibbons v. Ogden*, 22 U.S. (9 Wheat.) 1 (1824)).

[75] BARABÁSI, *supra* note 8, at 222–23.

[76] This comment seems prompted by the proximity of the second edition to the September 11, 2001 attacks.

[77] Although Yochai Benkler mentions network science in his excellent book on networks, the points made are more relevant to general regulation, and not specifically to copyright law. *See* YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 241–61 (2006).

[78] For more about the sharing practices in the early Internet, see ANDRÉS GUADAMUZ

technological utopia that had created an explosion in online content had also initiated a pirate's paradise.

In the late 1990s and early twenty-first century, the efforts to provide a legal framework that would accommodate the international nature of the online world were put in place. The 1996 World Intellectual Property Organization (WIPO) Copyright Treaty, the 1998 Digital Millennium Copyright Act (DMCA), and the 2001 Information Society Directive,[79] are just some of the legislative and international attempts to ensure that the new technology would not go unanswered. The economic importance of curbing widespread online copying would necessitate a broader view of copyright legislation. The objective of regulation proved to be twofold. First, emphasis was placed on enforcement, as exemplified by the often draconian measures of the DMCA.[80] Secondly, it was recognized that the technical nature of cyberspace would similarly necessitate technological solutions, so the protection of such protection measures would play a big part in this topic.[81]

Despite these and other regulatory attempts,[82] the Internet remains a very difficult territory for the enforcement of copyright. While it could be said that copyright legislation has never been stronger, and the enforcement by the copyright industry is relentless,[83] widespread copying of works online still takes place at levels that are increasing. Napster came into existence in 1999 and, at its height, it is believed that it had close to nineteen million users.[84] When Napster was successfully brought down by litigation,[85] new peer-to-peer ("P2P") technologies and applications such as Gnutella, Grokster, Kazaa and BitTorrent[86] managed to maintain and even increase illegal copying online. Despite a new barrage of increasingly successful litigation,[87] P2P traffic online has only continued to grow. Depending on the methods used to measure Internet transfer, it has been said that P2P transactions can hit as high as eighty percent of all recordable online traffic.[88] Not only is

GONZALEZ, COPYRIGHT IN CYBERSPACE: BUILDING FENCES ON THE INTERNET 17–28 (2002), http://papers.ssrn.com/sol3/papers.cfm?ab
stract_id=595362#PaperDownload.

[79] Council Directive 2001/29, 2001 O.J. (L 167) 10 (EC), *available at* http://www.ebu.ch/departments/legal/pdf/leg_ref_ec_directive_copyright_infosociety_220501.pdf.

[80] JESSICA LITMAN, DIGITAL COPYRIGHT: PROTECTING INTELLECTUAL PROPERTY ON THE INTERNET 166–170 (2001).

[81] Severine Dusollier, *Electrifying the Fence: The Legal Protection of Technological Measures for Protecting Copyright*, 21 EUR. INTELL. PROP. REV. 285, 285 (1999).

[82] *See also* Performances and Phonograms Treaty, Dec. 20, 1996, 36 I.L.M. 76; No Electric Theft (NET) Act of 1997, Pub. L. No. 105-147, 111 Stat. 2678 (codified as amended in scattered sections of 18 U.S.C.); Council Directive 2004/48, 2004 O.J. (L 157) 45 (EC) (describing the Directive of the European Parliament on the enforcement of intellectual property rights.

[83] Brett J. Miller, Comment, *The War Against Free Music: How the RIAA Should Stop Worrying and Learn to Love the MP3*, 82 U. DET. MERCY L. REV. 303, 309–11 (2005).

[84] This is a conservative estimate. *See* MICHAEL A. EINHORN, MEDIA, TECHNOLOGY AND COPYRIGHT: INTEGRATING LAW AND ECONOMICS 87–88 (2004) (discussing the drop in user base from 18.7 million to 150,000 people after Napster installed new security software to protect against copyright infringement).

[85] A&M Records, Inc. v. Napster, Inc., 284 F.3d 1091, 1098–99 (9th Cir. 2002).

[86] Bob Rietjens, *Give and Ye Shall Receive! The Copyright Implications of BitTorrent*, 2 SCRIPT-ED 327, 328–29 (2005) (describing the process of how BitTorrent functions as a P2P file distribution tool).

[87] *See, e.g.*, Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 936–37 (2005) (holding that "one who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties"); *In re* Aimster Copyright Litig., 86 F.App'x 984, 984 (7th Cir. 2004) (affirming an order of the district court requiring the defendant to pay a $5,000 fine and more than $100,000 in attorneys' fees for refusing "either to block infringing uses of his service or to shut the service down").

[88] Andrew Parker, Presentation: The True Picture of Peer-to-Peer Filesharing (2005),

this figure extremely high, but users of P2P networks shift from one application to another without signs of decline.[89]

What then can be done to enforce copyright on the Internet? Perhaps it is an impossible task, and there may be credence to an argument that cyberspace cannot be regulated.[90] Perhaps the copyright industries may have to realize that they must operate in a thoroughly new technological environment and, therefore, change the way in which they conduct business. The recent growth in "legal" downloads through successful sites like iTunes, and the potential of some P2P technologies for lawful uses,[91] may very well prove that a new era has arrived, and that cyberspace has indeed changed copyright forever.

But perhaps the problem with the regulation and enforcement of cyberspace responds to a basic lack of understanding of how this large network operates in reality. How do P2P networks grow? Who sustains them? Can we map infringement online? If we understand the Web better, then perhaps new regulation could respond to its technological challenges in a more efficient manner.

Network science may provide solutions for content owners. As mentioned earlier, one of the main characteristics of scale-free networks is that they are robust and stable.[92] As most of the links in a network are not hubs, an attack on one node is not likely to affect the entire network. The question then must be asked of whether P2P networks are scale-free, and therefore have inherited the characteristic reliability and robustness of such systems. After all, peer-to-peer networks have managed to survive, even after the forceful litigation experienced so far against the makers and hosts of clients and websites dedicated to infringing file sharing. This could provide evidence that P2P networks are indeed displaying the stability of scale-free architectures. In fact, these networks are so resilient to litigation that even the damning result against Sharman Studios, the makers of P2P client Kazaa, has not been enough to shut down their network.[93] Moreover, the network operates independently from the company that created it. But is the strength of P2P caused by the fact that it is a scale-free network, or because of other reasons?

When researchers have looked at P2P networks using the analytical tools described in earlier sections, they have found that they do indeed display scale-free characteristics, which would account for their strength. A study into the Gnutella P2P network found some inherent vulnerabilities, but concluded that:

> There are two mechanisms that cause the formation of scale-free topologies. First, networks expand continuously by the addition of new vertices, and second, new vertices attach preferentially to vertices that are already well connected. In Gnutella, the first

available at http://www.cachelogic.com/home/pages/studies/2004_01.php.

[89] *Id.*

[90] *See*, *e.g.*, Bill Thompson, *The Democratic Republic of Cyberspace?*, OPEN DEMOCRACY, Sept. 14, 2005, at 2–3 (2005), http://www.opendemocracy.net/content/articles/PDF/2832.pdf (arguing that the internet should not be governed by national governments, but instead by "a true deliberative forum that takes full advantage of the affordances of the internet itself and extends membership to all who wish to engage").

[91] *See, e.g.,* WOLFGANG NEDJL ET AL., EDUTELLA: A P2P NETWORKING INFRASTRUCTURE BASED ON RDF (2002), http://edutella.jxta.org/reports/edutella-whitepaper.pdf (discussing the use of P2P networks in the educational context).

[92] *See* Frank Kargl et al., *Protecting Web Servers from Distributed Denial of Service Attacks* (2001), http://medien.informatik.uni-ulm.de/forschung/publikationen/www10_01.pdf.

[93] Fasttrack, the Kazaa P2P network, was still running at the time of the writing of this paper.

> mechanism can be seen by the fact that new nodes are continuously entering and leaving the system, meaning the topology is undergoing constant change and growth. The second mechanism can be seen by the fact that there are only a few hosts that clients initially connect to. . . . Hence, the topology of the Gnutella network is scale-free because of its adherence to these two mechanisms.[94]

In another study, researchers charted some path-length in P2P networks in order to find out if they presented the small world phenomenon.[95] They discovered some power law behavior, but were surprised that links tended to cluster more than would otherwise be expected in a scale-free topology. The researchers then created their own P2P network, because they guessed that their results were being skewed by the efficiency of website search engines. The resulting link distribution between nodes in the network corresponded to power laws. Further research into the topic tends to corroborate these findings, and serves as a good indication that P2P networks are scale-free.[96]

If P2P networks are scale-free, it would be very bad news for copyright owners. Widespread file sharing online may not have the economic effects that are advertised by the industry,[97] but there must still be some effect.[98] Random attacks to client software, developers, and even to random users, will not bring down entire networks. This would mean that the industry may have to learn to live with the existence of P2P and change its business strategy to accommodate for their resilience, perhaps by offering better and improved services, as has been the experience with services such as iTunes and Rhapsody.[99] From latest figures and news, it is clear that random litigation against users has had no effect on downloads.[100]

However, enforcement online may still be possible. While scale-free networks are very resilient, they also can be the subject of catastrophic failures when a vital super-hub in the system collapses. As mentioned already, there may be situations where the super-node may produce cascading breakdowns in hubs and nodes directly connected to it, which could eventually bring down a network. While the Internet is very reliable, researchers calculate that an attack on five to fifteen percent of important hubs at the same time could eventually bring it down.[101] As P2P networks are scale-free,

---

[94] PEDRAM KEYANI ET AL., PEER PRESSURE: DISTRIBUTED RECOVERY FROM ATTACKS IN PEER-TO-PEER SYSTEMS 307 (2002), http://www.cs.cmu.edu/~pkeyani/publications/peerpressure.pdf.

[95] Mujtaba Khambatti et al., *Structuring Peer-to-Peer Networks Using Interest-Based Communities*, *in* DATABASES, INFORMATION SYSTEMS, AND PEER-TO-PEER COMPUTING 48 (Karl Aberer et al. eds., 2003), *available at* http://books.google.com (input the following search: "structuring peer to peer networks using internet based communities").

[96] *See* Lada A. Adamic & Bernardo A. Huberman, *Zipf's Law and the Internet*, 3 GLOTTOMETRICS 143, 147–48 (2002); Stefan Saroiu et al., *A Measurement Study of Peer-to-Peer File Sharing Systems*, http://www.cs.washington.edu/homes/gribble/papers/mmcn.pdf.

[97] Felix Oberholzer & Koleman Strumpf, *The Effect of File Sharing on Record Sales: An Empirical Analysis*, 115 J. POL. ECON. 1. (2007),.

[98] Rafael Rob & Joel Waldfogel, *Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students*, 49 J.L. & ECON. 29, 30, 60 (2006).

[99] It is calculated that in the United States alone, digital downloads of music, movies, and television shows amounted to $2.39 billion in 2006. *See* eMarketer.com, *Digital Downloading: Music, Movies and TV*, http://www.emarketer.com/Reports/All/Em_downloads_jan07.aspx?src=report2_home (last visited Mar. 29, 2007).

[100] *File-Sharing 'Not Cut By Courts'*, BBC NEWS, Jan. 20, 2006, http://news.bbc.co.uk/1/hi/entertainment/4627368.stm.

[101] BARABÁSI, *supra* note 8, at 118.

they are both robust and vulnerable. They are generally resistant to random attacks against nodes, such as the many instances in which the music and movie industries have been bringing infringement suits against P2P network users.[102] But what if the hubs in a specific network are identified? Then the enforcement strategy from copyright owners would be simple: attack the hubs and the network may be brought down.

Still, this strategy rests on the assumption that hubs can be identified, which may not always be easy to accomplish.[103] There may be some privacy implications in the release of detailed consumer data on the part of Internet Service Providers (ISPs); Canadian courts have already dismissed actions brought by the music industry when attempting to uncover the identity of file-sharers.[104] Some P2P clients and networks make it difficult to obtain individual information about the users, so identifying the hubs could also prove futile. Some of the most experienced hubs could also be masking their identity through software, or by connecting through virtual private networks (VPNs).[105]

Another particular vulnerability for P2P systems is that they seem to be remarkably prone to computer virus epidemics. According to Adamic and Huberman:

> Finally, it has been shown that scale-free networks are more susceptible to viruses than networks with a more even degree distribution. Namely, a virus spreading in a random network needs to surpass a threshold of infectiousness in order not to die out. However, if the network has a Zipf degree distribution, the virus can persist in the network indefinitely, no matter what level of its infectiousness.[106]

Despite the above considerations, one should take into account that there are legitimate uses for P2P networks, and that attacking one indiscriminately may affect some valid transactions. When dealing with copyright cases, the courts have already been reluctant to indict entire technologies if it can be assumed that they have significant non-infringing uses.[107] Nevertheless, in the case of P2P clients, the *raison d'être* for the technology is to promote infringement of copyright, as was demonstrated in *Grokster*, making them likely targets for litigation.[108]

While P2P clients have proved to be legally vulnerable, and could prove to be architecturally susceptible to attacks as well, other P2P systems may prove to be more resilient. As mentioned earlier, BitTorrent is the most popular sharing method in the world. BitTorrent seems to present similar characteristics to scale-free

---

[102] *See* Alice Kao, *RIAA v. Verizon: Applying the Subpoena Provision of the DMCA*, 19 BERKELEY TECH. L.J. 405, 406 (2004).

[103] *See* Jordana Boag, *The Battle of Piracy Versus Privacy: How the Recording Industry Association of America (RIAA) Is Using the Digital Millennium Copyright Act (DMCA) As Its Weapon Against Internet Users' Privacy Rights*, 41 CAL. W. L. REV. 241, 272 (2004).

[104] For more about this case, see Ian Kerr & Alex Cameron, *NYMITY, P2P & ISPs: Lessons from BMG Canada Inc. v. John Doe* (2005), *available at* http://iankerr.ca/files/Kerr-Cameron-NymityP2PISPs.pdf.

[105] A VPN is a private communications network used by an individual to communicate confidentially over a non-private network such as the Internet. The VPN traffic appears to come from the VPN server's domain and not from the actual location.

[106] Adamic & Huberman, *supra* note 96, at 148 (citation omitted).

[107] In *Sony Corp. of Am. v. Universal City Studios, Inc.* 464 U.S. 417 (1984), and in a lesser degree some of the Justices found so in *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005).

[108] *Metro-Goldwyn-Mayer Studios, Inc.*, 545 U.S. at 913.

topologies,[109] but the way in which it operates could prove to make it free of collapse.

BitTorrent is a communications protocol that distributes file-sharing amongst users with an entire copy of the whole (seeds), and/or amongst users with incomplete versions of the whole (peers). The information of who is sharing the files at any given time is distributed through a torrent file which connects to a tracking website, and allocates resources accordingly.[110] The distributed nature of the network makes it incredibly resilient,[111] and it also means that single attacks to hubs may eliminate one torrent file from the Internet, but many other copies can still remain. BitTorrent does rely heavily on tracker servers and search websites which may host torrent files, but because the actual infringing materials are not hosted in those sites, owners may have more difficulty getting rid of these.[112] There are also some legal ambiguities with the torrent system. The fact that BitTorrent is exceptionally useful in sharing bandwidth amongst all seeds and peers means that Bittorrent has become the protocol of choice to download licensed copies of large files, particularly software distributions and video.[113] This would make it immune from Grokster-like legal challenges.

Despite these questions, the potential for designing adequate, proportionate, and targeted enforcement strategies through the understanding of Internet architecture should not be neglected. When one reads about court orders against ISPs searching for the identity of P2P network users, one must wonder if the copyright industry knows about power laws already.[114]

Finally, better understanding of the way in which pages link to one another could also be used in order to determine potential risk of infringing practices found online. At present, calculations by the copyright industries of their monetary loss due to infringement are done with unsupported and/or exaggerated assumptions about consumer behavior.[115] By measuring accurately the status of an infringing node within the mesh of links that compose a scale-free network, it would be possible to calculate more accurately the actual loss suffered by the copyright owner.

B. Policy: The Long Tail

Enforcement and practical solutions are valuable, but the author believes that the true potential of network science with regard to copyright is in helping to shape policy. There seems to be a growing trend in intellectual property policy to draft future strategies based on evidence.[116] There are two recent successful examples of

[109] *See* M. IZAL ET AL., DISSECTING BITTORRENT: FIVE MONTHS IN A TORRENT'S LIFETIME, § 4, http://www.pam2004.org/papers/148.pdf (last visited Mar. 29, 2007).

[110] BitTorrent.org, For Developers, Protocol Specification (2006), http://www.bittorrent.org/protocol.html (last visited Mar. 29, 2007).

[111] ROBUSTNESS OF THE BITTORRENT PROTOCOL, § 5 http://mnl.cs.stonybrook.edu/home/karthik/BitTorrent/Robustness_of_BT.doc (last visited Mar. 29, 2006).

[112] *See* Rietjens, *supra* note 86, at 329.

[113] Examples of these are numerous. Some examples include Linux distributions, shareware, game demos, video podcasts, and etcetera.

[114] *ISPs Ordered to Reveal Software File-Sharers*, OUT-LAW NEWS, Jan. 31, 2006, http://www.out-law.com/page-6586.

[115] Oberholzer & Strumpf, *supra* note 97, at 2–3.

[116] One of the most outspoken supporters of this approach is Professor James Boyle. *See generally* James Boyle, *James Boyle: A Natural Experiment*, FIN. TIMES, Nov. 22, 2004, *available at* http://www.ft.com/cms/s/4cd4941e-3cab-11d9-bb7b-00000e2511c8.html (posing the question of how "we decide the ground-rules of the information age" and discussing the impact of adopting the "database right").

evidence-based policymaking in Europe. The first was the considerable public consultation process and research going into the discussion of the European Directive on Computer Implemented Inventions, which resulted in the eventual demise of the proposal.[117] The other example has been the Gowers Review of Intellectual Property,[118] which has made a big point of putting evidence before the interests of powerful lobbying groups.

If this trend is to continue, perhaps future policymakers will look at some of the results obtained from research of the copyright industries, and will shape their policy accordingly. But what does the research tell us?

The first effect that could be gleaned from current research, is that copyright industries seem to display Pareto distributions of wealth, exemplified by the often-commented phenomenon that most copyright earnings go to a comparatively small number of people.[119] The end result of the existence of a power law with regard to earnings, profits, and distribution of royalties may very well mean that most creators cannot expect to make a living from copyright, and only a minority of works will be successful.[120] This fact alone could be used when designing policies that may have larger effects on the public. When looking at the impact of new legislation, policymakers should take into consideration that creator-friendly policies will be more likely to benefit a minority of stakeholders, and should wonder if the larger sector of the public would benefit from such changes.

It seems like Pareto distributions are almost inevitable when it comes to measuring income, as results are skewed to allow for inequalities that seem to be almost inherent to such an analysis. When looking at data from the copyright industries, this is replicated in almost any field. From music to movies,[121] the story is the same: large amount of sales accumulate at the head of the graph, while smaller sales tend to tail off into the distance. In statistics, the resulting graph is sometimes called a "long-tailed distribution", and it has spawned an explanation tailor-made for network science, that of the Long Tail. In the word of its creator:

> The theory of the Long Tail is that our culture and economy is increasingly shifting away from a focus on a relatively small number of "hits" (mainstream products and markets) at the head of the demand curve and toward a huge number of niches in the tail. As the costs of production and distribution fall, especially online, there is now less need to lump

[117] *Commission Proposal for a Director of the European Parliament and of the Council, on the Patentability of Computer-implemented Inventions*, COM, Feb. 20, 2002, at 92; *see also* Andrés Guadamuz González, *The Software Patent Debate*, 1 J. INTELL. PROP. LAW & PRACT. 196 (2006).

[118] ANDREW GOWERS, GOWERS REVIEW OF INTELLECTUAL PROPERTY (2006).

[119] *See* RUTH TOWSE, CREATIVITY, INCENTIVE AND REWARD: AN ECONOMIC ANALYSIS OF COPYRIGHT AND CULTURE IN THE INFORMATION AGE 80–86 (2001) (describing the "superstar theory" which states that a few "superstars" who make exorbitant amounts of money skew the distribution of earnings in the music industry, and further explains the search and information costs in the market for singers); RUTH TOWSE, THE ECONOMICS OF COPYRIGHT: DEVELOPMENTS IN RESEARCH AND ANALYSIS 66, 68–69 (Wendy J. Gordon & Richard Watt eds., 2003) (stating that there is demand to concentrate the wealth of artists amongst a "few superstars").

[120] For example, in the United Kingdom, the most effective collecting society is the Mechanical-Copyright Protection Society (MCPS), which has more than 18,000 members and distributed in 2004 £219 million GBP among its members. Even if those profits were distributed equally, the average would be approximately £11,000 GBP. *See* The MCPS-PRS Alliance, About Us, http://www.mcps-prs-alliance.co.uk/aboutus (last visited Mar. 29, 2007).

[121] Newman, *supra* note 13, at 325.

products and consumers into one-size-fits-all containers.[122]

While the Long Tail does indeed respond to traditional Pareto distribution expectations, there is a surprising addendum when one looks at how sales charts behave when one adds Internet data into the equation. In traditional brick-and-mortar creative industries, the retail sector is specifically designed to respond to Pareto inequalities. Hits are given prevalence in shelf space all over music stores, bookshops, or DVD rental locales;[123] however, something strange is happening to these inequalities online. Electronic retailers still experience the occurrence of few massive hits and a long tail of less fortunate sellers, but when you factor out the need for limited shelf space, the tail keeps going, and never seems to disappear.[124]

Anderson offers several examples that help to explain this remarkable find: retail giant Wal-Mart shelves approximately 55,000 tracks in an average store, while digital music service Rhapsody has 1.5 million tracks; the remarkable fact is that Rhapsody's entire inventory has sold at least one copy.[125] In e-commerce giant Amazon, one third of total sales come from books that are outside of the top 100,000 list, and 57 percent of all book sales come from titles that are not stored in high-street book retailers.[126] The Long Tail therefore recognises that traditional media responds to power laws as profits go to a small cluster of entities. However, the Internet has provided a varied number of opportunities for those who did not have a chance to profit previously.

This leads to another counter-intuitive result from looking at the data emanating from the Long Tail which could have tremendous policy effects. A study into P2P file-sharing has unearthed the fact that sharing does indeed seem to affect music sales from top earners.[127] Blackburn conducted research trying to ascertain what would be the effect for music sales of a reduction of file-sharing volumes by thirty percent. For top earners at the head, the result was marked by an increase in sales; however, for those with minimum sales, decreasing file-sharing actually had a negative impact in sales.[128] If this data is accurate, then it could be said that P2P is good for the tail, but bad for the head.

A more tangential role in which network science may assist copyright policy is in gathering data and providing useful descriptions of the increasing role played by users in content creation in cyberspace; this is known as User Generated Content (UGC).[129] The UGC revolution is exemplified by the emergent popularity of blogging, mo-blogging, video blogging, podcasting, social networking, viral video and Wikipedia. The Internet is witnessing an explosion in content of momentous proportions. Most of this content inhabits the Long Tail: in the age of peer production and instant communications, the user has become a potential

[122] Chris Anderson, Long Tail FAQ, http://www.thelongtail.com/about.html (last visited Feb. 12, 2007).
[123] CHRIS ANDERSON, THE LONG TAIL: WHY THE FUTURE OF BUSINESS IS SELLING LESS OF MORE 19, 22 (2006).
[124] *Id.* at 19–23.
[125] *Id.*
[126] *Id.* at 23.
[127] *See generally* David Blackburn, *On-line Piracy and Recorded Music Sales* 1 (2004), *available at* http://www.katallaxi.se/grejer/blackburn/blackburn_fs.pdf.
[128] *Id.* at 45–46.
[129] Benkler calls it "peer production." *See* BENKLER, *supra* note 77, at 89.

publisher. The dissemination of information witnessed throughout the Web makes more sense when it is viewed through the lens of network science. If economic incentives are largely irrelevant for a good number of the people involved in online publishing, then what is the currency of cyberspace? Some have proposed that information was the currency of cyberspace, where it could be exchanged for more information.[130] But it now seems that the real currency online are links. Cyberspace is a network: nodes link to other nodes, and the more links you have the more likely you are to grow. What should then be the strategy of a website? To collect more links, no matter what.

The relevance for network science here is that because of structural issues, the Internet may favor the prevalence of UGC over commercial creative works because it seems to encourage the creative process, and its ulterior dissemination to the entire Web. Policymakers should take this into consideration when looking at ways to regulate copyright in the digital domain. There is growing evidence that the vast majority of copyright owners are located in the peer production sector, and therefore, policy should reflect that. It would be useful if policy was no longer designed with the idealized struggling creator in mind. More often than not, the creator will be a hobbyist, never expecting a monetary return for her troubles.[131] Nevertheless, it is easy to make this point while forgetting that while peer production is on the increase, readership of such content may not be. Just because something is online does not mean that it has an audience.[132] Similarly, "traditional" offline world ideas of quality and peer-review still apply to the online environment.

C. Dissemination: Open Licensing

The third and last suggested area of exploration for copyright-related research is to use network science to chart the dissemination and propagation of licensed work throughout the Web, particularly "open" content distributed through open source software or open content license.[133]

It is now clear that websites respond to power laws, which would probably indicate that the currency of information is unequal in online environments, as the hubs dominate the landscape. The inequality prevalent in scale-free networks has dissemination implications if we see copyright as a mere economic right. If that is the case, then the phenomena described by network science have economic relevance for the collection of profits, as explained in the previous section. Emphasis on economic value will serve to dispel the utopian myths of the Internet as a democracy of information where data is the currency.[134] There would be an inherent

---

[130] Hillary Bays & Miranda Mowbray, *Cookies, Gift-Giving, and the Internet*, FIRST MONDAY, Nov. 1999, http://firstmonday.org/issues/issue4_11/bays/index.html.

[131] The death of the traditional view of the author has already been explored. *See generally* ROSEMARY J. COOMBE, THE CULTURAL LIFE OF INTELLECTUAL PROPERTIES: AUTHORSHIP, APPROPRIATION, AND THE LAW (1998); THE CONSTRUCTION OF AUTHORSHIP: TEXTUAL APPROPRIATION IN LAW AND LITERATURE (Martha Woodmansee & Peter Jaszi eds., 1994).

[132] *See* BENKLER, *supra* note 77, at 243–45.

[133] It will be assumed that the reader is familiar with the terms. For some more information about the terminology, see generally Andrés Guadamuz González, *Open Science: Open Source Licences in Scientific Research*, 7 N.C. J. L. & TECH. 321 (2006); Andrés Guadamuz González, *Viral Contracts or Unenforceable Documents? Contractual Validity of Copyleft Licenses*, 26 EUR. INTELL. PROP. REV. 331 (2004) [hereinafter *Viral Contracts*].

[134] Some of these ideas are expressed in, LAWRENCE LESSIG, THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD 241–44 (2001).

inequality on the nature of information, everybody is free to participate in the online environment, but only the works at the head of the long tail will be of importance. But copyright is not only about economics. The popular acceptance of Pareto distributions exemplified by the 80/20 rule shows us that people are still willing to create without hope of remuneration. There is indeed an economic incentive provided by copyright, but this incentive is nebulous; many people are happy to produce works subject to copyright protection for all sorts of reasons. There is something to be said about those who willingly inhabit the long tail.

The argument explored in the previous section is that the vast majority of creators produce works without clear expectation of economic reward, which can explain the rise of non-proprietary production.[135] Open source software, free software, free culture, open access, and many other non-proprietary forms of development and distribution are protected by copyright, but they tend to be released under permissive "some rights reserved" licenses such as Creative Commons.[136]

But the desire to disseminate copyright works through open licenses does not only respond to Pareto distributions and Long Tail inequalities. The very structure of open source software development displays scale-free topologies.[137] When studying how open source projects operate and become organized, researchers have found that the pattern of individual interactions between programmers is not random, but is instead scale-free.[138] Open source software developers tend to join a collective, and contribute source code into the project. It seems like a few key developers act as network hubs, holding together the project. Unsurprisingly, other open content projects—such as Wikipedia—seem to display the same reliance on super-hubs.[139]

While knowing how open source projects operate and organize into complex networks may be interesting from a social and economic perspective, there are other implications of network science that are more relevant to the legal field, and that is in the area of copyleft licensing. Copyleft is a legal concept drafted by programmer Richard Stallman, which allows the further distribution and modification of software released under the terms of a copyleft licence, but it does not allow proprietary developers to come and "close" the released software by not offering access to the source code.[140]

The GNU General Public License (GPL)[141] is the most important copyleft license out there.[142] One of the most important clauses

---

[135] *See Viral Contracts*, *supra* note 133, at 338–39.

[136] For more about Creative Commons and other licenses, see AHRC RESEARCH CENTRE FOR STUDIES IN INTELLECTUAL PROPERTY AND TECHNOLOGY, UNIVERSITY OF EDINBURGH, THE COMMON INFORMATION ENVIRONMENT AND CREATIVE COMMONS: FINAL REPORT TO THE COMMON INFORMATION ENVIRONMENT MEMBERS OF A STUDY ON THE APPLICABILITY OF CREATIVE COMMONS LICENSES (2005), *available at* http://talk.talis.com/archives/CIE_CC_Final_Report.pdf.

[137] *See* Georg von Krogh et al., *Community, Joining, and Specialization in Open Source Software Innovation: A Case Study*, 32 RES. POL'Y 1217, 1226–27, 1229 (2003).

[138] KEVIN CROWSTON & JAMES HOWISON, THE SOCIAL STRUCTURE OF FREE AND OPEN SOURCE SOFTWARE DEVELOPMENT TEAMS (2004), *available at* http://freesoftware.mit.edu/papers/crowstonhowison.pdf.

[139] JAKOB VOSS, MEASURING WIKIPEDIA 9–10 (2005), http://eprints.rclis.org/archive/00003610/01/MeasuringWikipedia2005.pdf.

[140] *See* GLYN MOODY, REBEL CODE: THE INSIDE STORY OF LINUX AND THE OPEN SOURCE REVOLUTION 26–29 (2002).

[141] GNU General Public License, http://www.gnu.org/copyleft/gpl.html (last visited Mar. 29, 2007).

[142] At the time of writing, 67 percent of all projects listed in the SourceForge open source repository are released under the GPL. *See* SourceForge.net: Software Map,

included in the GPL is the copyleft clause, which sets restrictions against using the software in proprietary manners. The section reads:

> 2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions: . . . b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.[143]

What this means is that any software developed by using the open source code of the copyleft program must not charge for the derivative product, and most importantly, must ensure that the GPL is transferred to further users of the derivative software. This type of licence has been aptly named a "viral contract" as the contractual obligations contained are passed through a chain of distribution to other contracting parties.[144] The GPL therefore spreads in viral form, as the licensee must include the terms of the GPL in any subsequent derivative work they produce. Those subsequent licensees will be under the obligation to license their derivatives with the same obligations in place, and so on and so forth.

This is where open network science may become useful. One of the effects of copyleft licensing is that it may generate long chains of distributions in which software projects may fork into others, with the originating source code becoming entangled in licensing webs.[145] Under the GPL, the copyleft aspects only apply to derivative works that are distributed to the public. This simple rule could generally provide clear-cut cases in which the GPL would apply and where it would not. For example, software that simply interacts with other GPL software does not suffer from copyleft restrictions. However, the current situation is that the copyleft clause contained in the GPL is now being re-drafted in order to provide added protection. The draft of the new version of the GPL is under discussion at the time of writing, but it is clear that it could prove problematic.[146] The current draft reads:

> b) You must license the entire modified work, as a whole, under this License to anyone who comes into possession of a copy. This License must apply, unmodified except as permitted by section 7 below, to the whole of the work. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have

---

http://sourceforge.net/softwaremap (last visited Mar. 29, 2007). Out of 73,978 projects in the SourceForge repository, 50,013 were GPL software. *Id.*

143 GNU General Public License, *supra* note 141.

144 Margaret Jane Radin, *Humans, Computers, and Binding Commitment*, 75 IND. L. J. 1125 (2000). Some people in the Free Software community do not like the term "viral", but it has become more prevalent in the literature. For example, see THORSTEN WICHMANN, FIRMS' OPEN SOURCE ACTIVITIES: MOTIVATIONS AND POLICY IMPLICATIONS, REPORT FOR THE FREE/LIBRE OPEN SOURCE SOFTWARE: SURVEY AND STUDY (2002), http://strategicadvice.net/floss.pdf.

145 Robert W. Gomulkiewicz, *General Public License 3.0: Hacking the Free Software Movement's Constitution*, 42 HOUS. L. REV. 1015, 1028–29 (2005).

146 For more about this issue, see Andres Guadamuz, *GNU General Public License v3: A Legal Analysis*, 3 SCRIPT-ED 130 (2006).

separately received it.[147]

This leaves less clear whether unrelated code included with the licensed work would have to be licensed under the GPL. If this is the case, it may become extremely useful to chart potential clusters of GPL software through the application of network science in order to provide a clear mapping of potential licensing pitfalls. Having a good idea of the small world paths linking software may also help to solve potentially difficult questions about code ownership that have been seen in some cases.[148]

V. CONCLUSION

This paper has been presented in various stages to diverse audiences.[149] The last slide of the presentation features a wonderful picture of the footpaths in a public park in Stuttgart University.[150] The picture shows the designed path by whoever built the space, a nice crossing X through the roughly square lawn. However, one can clearly see another path in the picture, one that was not designed. This path has been made by people walking from one building to another in a direct line, which does not follow the official pathway. This exemplifies nicely what the present article is trying to achieve. We may plot paths through cyberspace; we may attempt to regulate the space in various ways. But is this regulation really considering the paths that will be chosen almost inevitably by the inhabitants of the new space? Network science provides a descriptive tool to make better decisions when building the paths.

One such path is the role of copyright regulation on the Internet. The better understanding of how the global network operates may provide enhanced tools to regulate and enforce copyright law in cyberspace. The objective of the article has been to look at some of the corollaries arising from network science and apply them to three fields of copyright law, enforcement, policy and dissemination of works. Some of the issues highlighted are in initial form of research, or we may still need more data in order to properly suggest policy shifts. This has been an initial exploration in order to flag some of the important issues where the author believes there are further avenues of research for legal scholarship.

---

[147] GNU General Public Licenses, Discussion Draft of Version (2006), http://gplv3.fsf.org/comments/gplv3-draft-2.html.

[148] An example can be found in the ongoing case of *S.C.O. Group, Inc. v. I.B.M.*, No. 03-CV-294, 2005 WL 318784 (D. Utah 2005).

[149] To mixed results, ranging from outraged to enthusiastic.

[150] BALL, *supra* note 10, at 169.